Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application. Applicant has submitted a new complete claim set showing marked up claims with insertions indicated by underlining and deletions indicated by strikeouts and/or double bracketing.

Listing of Claims:

1. (Currently amended) In a database system, a sampling method for constructing a data structure based on the contents of a database comprising:

selecting~~a) gathering~~ an initial sample of data from the database, the initial sample of data including one or more subparts;

cross-validating a plurality of subparts of the ~~and creating a first data structure from said~~ initial data sample, the cross-validating associated with an error corresponding to a subpart;

sorting substantially simultaneously with the cross-validating the plurality of subparts to generate a plurality of cross-validation errors;

generating an estimated block size based on the sorting and cross-validating;

selecting an additional ~~b) gathering a second~~ sample of data, wherein the size of the selected additional sample of data corresponds to the generated estimated block size~~from the database~~;

merging the

~~c) determining an initial sufficiency of the data gathered from the database that is based on a comparison of the first data structure and the second sample of data; and~~

~~d) forming a resultant data structure by gathering an~~ additional sample of data with the ~~from the database and using the additional amount of data to form the~~

~~resultant data structure wherein the amount of data gathered in the additional sample is based on the~~ initial sample of data~~sufficiency determination~~.

2. (Currently amended) The method of claim 1 wherein the cross-validating includes cross-validating subparts of ~~resultant data structure is formed based on data gathered in~~ the initial sample of data that are of different sizes~~, the second sample and the additional sample~~.

3. (Currently amended) The method of claim 1 wherein the cross-validating and sorting are combined in a single step ~~first and resultant data structures are histograms~~.

4. (Currently amended) The method of claim ~~1~~ 3 wherein the single step includes:

dividing the initial sample of ~~and second~~ data into multiple subparts;

sorting and cross-validating the multiple subparts recursively~~samples are randomly retrieved block samples that form a first amount of data that is initially gathered and then divided in half to provide the initial and second data samples~~.

5. (Currently amended) The method of claim 4 wherein the single step further includes:

building a histogram for at least a first subpart and a second subpart of the initial sample of ~~and second~~ data;

testing the histogram of the first subpart against the second subpart to generate a cross-validation error estimate for a sample size corresponding to the initial sample of data ~~samples are sorted and used to form two histograms~~.

6. (Currently amended) The method of claim 5 further comprising reusing parts of the initial sample of data to generate different cross-validation ~~wherein an~~ error estimates, each of the cross-validation error estimates corresponding to an associated sample size~~metric of the two histograms are formed by cross correlating the contents of the two histograms to determine the initial sufficiency~~.

7. (Currently amended) The method of claim 6 wherein generating the estimated block size includes:

computing means of the different cross-validation error estimates for each of the associated ~~the initial and second data samples are further sub-divided to form sub-samples used to form other histograms of differing~~ sample sizes;

determining a best fit of the means of the different cross-validation error estimates;

estimating the block size based on the determined best fit~~that are cross correlated to find an error metric relating to said differing sample sizes~~.

8. (Currently amended) The method of claim ~~6~~ 7 wherein determining the best fit includes identifying a best fitting curve associated with the means of the different cross-validation ~~initial and second data samples are further sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an~~ error estimates~~metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database~~.

9. (Currently amended) The method of claim ~~4~~ 8 wherein identifying a best fitting curve includes:

generating the best fitting curve of the form $\Delta^2 = c/r$, wherein c is a constant, Δ^2 is an average squared cross-validation error observed for a given sample size, and r represents the given sample size;

estimating the block size based on the contant c ~~additionally comprising estimating distinct values of an attribute of the initial and second samples by eliminating records from the blocks that are duplicated within a given block and estimating distinct values by categorizing attributes as rarely or frequently occurring within the database~~.

10. (Original) A computer readable medium for performing computer instructions to implement the method of claim 1.

11. (Currently amended) A database system for constructing histograms based on sampling the contents of the database comprising:

a) a database management component that gathers block size data segments from the database which in aggregate form a first sample of data having a first size;

b) a histogram construction component that forms a first histogram from the first sample of data; and

c) a correlation component that cross-validates a plurality of subparts of the initial sample of data and sorts substantially simultaneously with the cross-validating the plurality of subparts to generate a plurality of cross-validation errors, ~~determines an initial sufficiency of the first sample of data gathered from the database based on a comparison of the first histogram and data from the first sample of data;~~

~~d)~~ wherein said database management component gathers an additional sample of data used by said histogram construction component in creating a resultant histogram corresponding to a combination of the additional sample and the initial sample of data, ~~and~~ the size of the additional sample being ~~is~~ based on the cross-validation errors ~~initial sufficiency determination~~.

12. (Currently amended) The system of claim 11 wherein the resultant histogram is formed by the histogram construction component based on data gathered in the first sample of data and the additional sample of data.

13. (Original) The system of claim 11 wherein the first sample of data and the additional sample of data are randomly retrieved block samples.

14. (Original) The system of claim 11 wherein histogram construction component sorts the data in said first sample of data as it constructs the first histogram.

15. (Currently amended) The system of claim 11 wherein the correlation component determines the cross-validation errors ~~an error metric~~ by cross correlating the contents of the first histogram with other data in said first sample of data to determine an~~the~~ initial sufficiency of the first sample of data gathered from the database.

16. (Currently amended) The system of claim 15 wherein the first sample of data is sub-divided to form ~~sub-samples~~the subparts used to form histograms of differing sizes that are cross correlated to find a cross-validation ~~an~~ error ~~metric~~ relating to said differing sample sizes.

17. (Currently amended) The system of claim 15 wherein the first sample of data is sub-divided to form additional ~~sub-samples~~subparts of smaller size that are used to form other histograms that are cross correlated for use in finding cross-validation erros ~~an error metric~~ relating to sample sizes for use in determining a size of the additional sample of data to gather from the database.

18. (Currently amended) In a database system, a sampling method for constructing a histogram based on the contents of a database comprising:

a) gathering an initial sample of data from the database and creating a histogram from said initial sample;

b) gathering a second sample of data from the database for comparison with said first histogram;

c) determining an initial sufficiency of the data gathered from the database that is based on a comparison of the second sample with the first histogram, including cross-validating and sorting a plurality of portions of the data substantially simultaneously; and

d) if the determination of initial sufficiency indicates the data in said initial and second samples is adequate to represent the database, combining the initial and second samples to form a resultant histogram, but if the determination of initial sufficiency indicates the initial and second samples are inadequate to represent the database, gathering an additional data sample to combine with the initial and second samples to form the resultant histogram wherein a size of the additional data sample is based on the initial sufficiency determination.

19. (Original) The method of claim 18 wherein the data is gathered in blocks from random storage locations within the database.

20. (Currently amended) In a database system, a system for constructing a data structure based on the contents of a database comprising:

a) means for gathering an initial sample of data from the database and creating a first data structure from said initial sample;

b) means for determining an initial sufficiency of the data gathered from the database that is based on a comparison of the first data structure and other data in the initial sample not used to create the first data structure, the comparison being based on cross-validating and substantially simultaneously sorting a plurality of portions of the data; and

c) means for forming a resultant data structure by gathering an additional sample of data from the database and using the additional amount of data to form the resultant data structure wherein the amount of data gathered in the additional sample is based on the initial sufficiency determination.

21. (Original) The system of claim 20 wherein the resultant data structure is formed based on data gathered in the initial sample and the additional sample.

22. (Original) The system of claim 21 wherein the first and resultant data structures are histograms.

23. (Original) The system of claim 20 wherein the initial data sample is made up of randomly retrieved block samples that form a first amount of data that is

divided in half to provide data to form the data structure and data to cross correlate against the first data structure.

24. (Original) The system of claim 23 wherein the initial data samples is sorted and used to form two histograms.

25. (Original) The system of claim 24 wherein an error metric of the two histograms are formed by cross correlating the contents of the two histograms to determine the initial sufficiency.

26. (Original) The system of claim 25 wherein the initial data sample is further sub-divided to form sub-samples used to form other histograms of differing sample sizes that are cross correlated to find an error metric relating to said differing sample sizes.

27. (Original) The system of claim 26 wherein the initial and second data samples are further sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an error metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database.

28. (Original) The system of claim 24 additionally comprising means for estimating distinct values of an attribute of the initial and second samples by eliminating records from the blocks that are duplicated within a given block and estimating distinct values by categorizing attributes as rarely or frequently occurring within the database.

29-31. (Canceled)

32. (Original) A computer readable medium for performing computer instructions to implement the method of claim 20.

33-34. (Canceled)